Science Education
Review Letters

Analyzing the FCI based on a Force and Motion Learning Progression

Irene Neumann[a*], **Gavin W. Fulmer**[b] **and Ling L. Liang**[c]
[a] Department of Physics Education, Ruhr-Universität Bochum, Bochum, Germany
[b] National Institute of Education, Singapore
[c] La Salle University, Philadelphia, PA, USA

[*] Corresponding author: neumann@physik.rub.de

Received 6th November 2012, Accepted 28th May 2013

Abstract
In this paper, we bring together a well-established and often-studied instrument assessing students' understanding about force and motion, the Force Concept Inventory (FCI), and a proposed description of how students develop in understanding of the force concept, the Force and Motion Learning Progression (FM-LP). We report on two phases of content analysis of the FCI and the FM-LP. In the first phase, findings indicate that 17 FCI items address aspects consistent with the FM-LP. In the second phase, our findings show that these 17 items have responses that can be coded to fit the levels of the proposed FM-LP. Implications for future research on both the FCI and the FM-LP are described.

Keywords
Learning Progressions, Force Concept Inventory, Force Concept, Assessment of Students' Understanding, Physics.

Introduction

Force is one of the central concepts in physics. It is used to describe the behavior of objects in electromagnetic or gravitational fields, for example. From the perspective of physics education, force is considered a core idea, which can be used to organize the more detailed elements of physical science content and to form the basis upon which new knowledge can be built (National Research Council [NRC], 2012). Much work has been done to identify students' conceptions about force (e.g., Alonzo & Steedle, 2009; see Duit, 2010, for an overview). Prior studies have shown inadequate conceptions being quite robust and persisting through physics education up to the university level (e.g., Jung & Wiesner, 1981; Schecker & Niedderer, 1985). As a consequence, physics teaching aims at promoting an adequate understanding of force. To achieve this aim also requires the development of instruments that can measure students' understandings and that draw upon current research on physics teaching and learning.

Theoretical Background

Recently, Alonzo and Steedle (2009) drew on the emerging and growing field of Learning Progressions (LPs) and proposed a "Force and Motion Learning Progression" (here abbreviated as FM-LP) describing the growth of understanding the force concept through instruction. Learning progressions are descriptions of the increasingly complex understand-

Published on 20/06/2013 on http://edoc.hu-berlin.de/serl

ings of a scientific concept that students can develop over multiple years of instruction (NRC, 2007; Duschl, Maeng, & Sezen, 2011). The concept of a learning progression (LP) carries with it both information about students' cognitive development and common alternative conceptions (e.g., Liu & Lesniak, 2006; Talanquer, 2009). It also has implications for the development of curriculum, pedagogical approaches, and assessments that can relate to and accommodate the different understanding over time (e.g., Duncan & Hmelo-Silver, 2009; Lee & Liu, 2010; Mohan, Chen, & Anderson, 2009; Wilson, 2009). LPs must connect to a specific concept or competence, such as matter (Stevens, Delgado, & Krajcik, 2010), carbon cycling (Mohan, Chen, & Anderson, 2009), celestial motion (Plummer & Krajcik, 2010), or argumentation (Berland & McNeill, 2010). LPs are inherently hypothetical in their initial development (e.g., NRC, 2012). Thus, their use as a conceptual basis for understanding students' learning over time and as anchors for curriculum or assessment requires iterative validation and refinement (Shavelson & Kurpius, 2012).

Alonzo and Steedle (2009) emphasize the ambiguity of LPs "in terms of what is being portrayed […] and in their breadth and grain size" (p. 391). Building on Wilson's (2008) considerations on "relationships between construct maps and learning progressions" (Alonzo & Steedle, 2009, p. 392), Alonzo and Steedle note that LPs of larger concepts might be made up of LPs addressing smaller aspects of the respective concept. In proposing the FM-LP, Alonzo and Steedle reviewed and built upon prior research on students' scientific and alternative conceptions of force and motion, and focused on one-dimensional forces and resulting motion. The FM-LP seeks to describe the growth of students' understanding across five levels ranging from "way off track" (level 0) up to an understanding of the relation between net force and acceleration (level 4). Each level includes a general description of students' typical thinking and respective explanations in the cases of force/no force and motion/no motion (see Alonzo & Steedle, 2009, pp. 403-405). Alonzo and Steedle also developed and applied an instrument to assess student understanding relative to the FM-LP and its respective levels.

As a core concept in physics, student understanding of the force concept has been extensively studied. For instance, the Force Concept Inventory (FCI; Hestenes, Wells, & Swackhamer, 1992), a well-established and widely used instrument assessing students' conceptions about force, has been examined and validated in many studies since its initial publication (e.g., Planinic, Ivanjek, & Susac, 2010; Savinainen & Scott, 2002; Stewart, Griffin, & Stewart, 2007; Wang & Bao, 2010). The FCI was constructed based on considerable prior research on students' conceptions and consists of 30 items, each of which poses a physical situation and consists of five response options—one that corresponds to the correct physical conception, and four others that correspond to common alternative conceptions. Though it was suggested that the FCI should be treated as a diagnostic instrument for qualitative understandings of students' thinking (e.g., Hestenes & Halloun, 1995), prior research has typically treated the FCI items as dichotomous (e.g., Liang, Fulmer, Majerich, Clevenstine, & Howanski, 2012; Planinic et al, 2010; Wang & Bao, 2010). However, this dichotomous scoring discards potentially important information on students' reasoning about the physical phenomenon that is contained in the response options (Briggs, Alonzo, Schwab & Wilson, 2006).

Making use of the information included in both correct responses *and* distractor options is the underlying idea of so-called Ordered Multiple-Choice (OMC) items (Briggs et al., 2006). OMC items' response options are designed to represent a particular level of understanding mapping on a model of students' development in thinking (Briggs et al., 2006). "Taken together, a student's responses to a set of OMC items permit an estimate of the student's level of understanding" (p. 43), that is OMC items combined with a cognitive model (e.g., an LP) provide a criterion-referenced diagnostic capability. The instrument used in Alonzo and Steedle (2009) to assess student understanding relative to the FM-LP was based on OMC items. Drawing on the use of OMC items and the application to the FM-LP, we analyze the FCI to make use of the information contained in the distractor response options to examine if, by treating the FCI items as OMC items, students' results on the FCI could be evaluated relative to a learning progression, thus providing a detailed description of students' thinking.

Though the FCI and FM-LP derive from different perspectives about student reasoning, there are certain commonalities between the two as both address students' conceptions of force and motion and as both build upon prior study of

Published on 20/06/2013 on http://edoc.hu-berlin.de/serl

students' reasoning about physical phenomena and related alternative conceptions and common errors. However, FM-LP is focused on only an aspect of the force concept, that is the relation of force and motion in one-dimensional situations (Alonzo & Steedle, 2006), whereas the FCI more broadly covers the force concept, e.g. by addressing each of Newton's laws of motion (Hestenes et al., 1992). As there is an overlap of addressed concepts in FM-LP and FCI, in this study we examine the relative fit between the well-established FCI and the FM-LP through the lens of the recently emerging field of LPs by posing the following research question: To what extent could the FM-LP serve as a rubric for scoring the FCI responses? Through such analysis, this study's findings can enable a more detailed diagnostic capability of the FCI, which takes into account the recent work on LPs.

Methodology

We conducted an expert rating study to answer the research question. The main focus guiding the rating was on the matching of the FCI items' response options to the FM-LP levels. The authors served as content analysts, drawing upon our expertise in physics and physics education (further referred to as raters A, B, C). The materials included in the expert rating study were all items from the FCI (revised version by Halloun, Hake, & Mosca, 1995) as well as the force and motion LP by Alonzo and Steedle (2009, pp. 403-405). The expert rating process included independent rating phases followed by group discussions to reach consensus on divergent ratings. We used Fleiss' (1971) κ index to identify the degree of agreement among the ratings. Fleiss' κ is an extension of Cohen's κ to multiple raters, without any weighting adjustment for the ratings. It provides an index of the degree of agreement among multiple raters. It ranges from 0 to 1, with 1 indicating perfect agreement among all raters. We also examined pairwise inter-coder agreement percentages as further substantiation of the findings.

Results

In the first step, we independently coded all 30 FCI item responses into the LP levels. We also noted any difficulties when coding particular items based on the LP levels. A group discussion revealed that these differences originated from some items addressing content aspects of the force concept that were not precisely addressed by the FM- LP. Therefore, in the second step, we independently rated all 30 items about the matching of the FM-LP content on the one hand, and the content addressed by the FCI item on the other. An overall agreement on the match between the FM-LP and the FCI items' content was quite high (Fleiss' κ =.91). Two authors had perfect agreement (A-B: 100%), and the other disagreed on only two of the 30 items (A-C, B-C: 93% agreement). After discussion, we reached consensus that 17 items addressed the FM-LP's content and would be retained. The remaining 13 items touched on other content aspects of the force concept which have not been included in the FM-LP: comparing the motion of bodies with different masses (2 items); Newton's third law (4 items); circular motion (4 items); quantitative view on velocity as a vector (1 item); and comparing velocities and accelerations using position data (2 items).

In the third step, we assigned each FCI response option to one level of the FM-LP for the 17 items retained from the previous step. We assigned the LP level that was deemed minimally necessary for a student to select a particular response option. For these 85 instances (five options per 17 items) the degree of agreement was determined but, due to a double coding by one author of one instance to two possible LP levels, data from only 84 instances were analyzed. The overall agreement was quite low (Fleiss' κ =.41). The percentages of agreement were 44% (B-C), 55% (A-B), and 66% (A-C). In order to moderate the rating, we then discussed four sample items (comprising 20 instances). Each author described the reasons for his or her decision on the level assignment. Then, the reasons were distributed among the authors and discussed. During this discussion, we reached a consensus on the level assignment for these 20 instances, and clarified the level interpretations (for details see Appendix). Note that we did not change the FM-LP itself. In the final step, the remaining 13 items covering the FM-LP's content were rated for a second time. Again, each of the items' responses was assigned to a specific LP level. The overall agreement was much improved compared to the rating before moderation (Fleiss' κ =.74; percentages of agreement were 77% [A-B], 79% [B-C], and 85% [A-C]). A subsequent discussion resulted in a consensus assignment of response options to LP levels among the experts (see Table 1).

Published on 20/06/2013 on http://edoc.hu-berlin.de/serl

Table 1. Final Consensus Coding of FCI Responses (A-E) According to the LP Levels (0-4).

Item No.	Key	A	B	C	D	E	Item No.	Key	A	B	C	D	E
Q01			n/a				Q16			n/a			
Q02			n/a				Q17*	B	3	4	0	3	0
Q03	C	3	3	4	1	2	Q18			n/a			
Q04			n/a				Q19			n/a			
Q05			n/a				Q20			n/a			
Q06			n/a				Q21	E	3	2	3	0	4
Q07			n/a				Q22*	B	3	4	0	3	0
Q08	B	2	4	2	2	2	Q23	B	2	4	2	2	0
Q09			n/a				Q24	A	4	0	3	0	3
Q10	A	4	0	3	2	3	Q25	C	0	0	4	3	0
Q11	D	1	2	2	4	0	Q26	E	3	3	0	3	4
Q12*	B	2	4	2	2	2	Q27	C	2	2	4	0	0
Q13	D	3	2	3	4	0	Q28			n/a			
Q14	D	0	1	2	4	2	Q29	B	1	4	2	3	2
Q15			n/a				Q30*	C	1	2	4	2	2

Note. Starred items were used for moderating during the coding process. Items marked “n/a” are coded as not having content match with the FM-LP.

Discussion

This paper aimed at bridging the extensive work on a well-established instrument measuring students’ conceptions about force (the FCI) and the emerging field of learning progressions that is “rapidly gaining popularity in the science education community” (Alonzo & Steedle, 2009, 391). An expert rating was used to view the response options of FCI items through the lens of learning progression research.

Summary of the findings

Our findings indicate that the FM-LP can serve as a rubric for scoring the FCI responses into polytomous categories, to a certain extent. Whereas the FM-LP is focused on only the relation between force and motion in one-dimensional cases, the FCI covers the force concept more broadly. The following aspects are covered by FCI only: (1) quantitative perspective on forces (comparing different masses); (2) the concept of action and reaction (Newton’s third law); (3) circular and two-dimensional motion; and (4) velocity as a vector; and (5) conceptual understanding of velocity and acceleration. So using the FM-LP for analysis of students’ thinking that is more detailed than simple dichotomous coding is only possible for a subset of FCI items.

Seventeen items were found to address similar aspects of force as the FM-LP. Among those items we developed a consensus assignment of item response options to FM-LP levels. Viewing FM-LP as a cognitive model according to Briggs et al. (2006) this FCI subset could therefore be used to measure students’ level of understanding on the LP, but only if taken together. From a measurement point of view a student’s response to a single FCI item would not reveal reliable results. The consensus assignment shows a complete, but unequal, covering of the FM-LP levels through the FCI response options. That is, LP levels 0 through 4 are assigned to 18, 5, 26, 19, and 17 responses, respectively. Clearly, some LP levels are measured by markedly fewer response options. The lack of balance among various LP levels is not surprising given that the FCI was not designed as an OMC instrument in the first place. Yet in general our findings show that FCI and FM-LP can be related to each other to a significant extent.

Implications for Future Research

The results of our study contribute to science education research in two ways: (1) informing research using the FCI; and (2) informing research on learning progressions on the force concept. Regarding the FCI, the findings highlight the importance of examining the different conceptual bases for the FCI items’ response options. Each item’s response options provide information not only regarding the students’ mastery of the force aspect measured by the item itself, but also about the alternative conceptions of force that the students hold. Scoring the FCI items dichotomously as right or wrong, therefore, destroys valuable information about students’ thinking. Future studies using the FCI items should consider viewing the FCI as OMC

Published on 20/06/2013 on http://edoc.hu-berlin.de/serl

items rather than dichotomous analysis, and the FM-LP or a similar coding scheme could be used for this purpose.

Additionally, while the FCI items' responses could be assigned a rating according to the FM-LP levels, the responses for any single item do not always cover all levels. This in part reflects the original goals for the FCI as a diagnostic rather than an LP-related instrument. In some cases, two response options for an item are assigned to the same level. In such cases, the items may not have any responses assigned to levels 1, 2, or 3, for example. In general, this would not be a problem. However, if one of the LP levels is underrepresented across the entire item set, the measurement error for this level could be quite significant because there would be inadequate evidence to estimate it. If a better fit between the FCI and the FM-LP were desired, then the FCI items would need to be refined so that more of the items had responses that better represented the LP levels. This would allow a more precise measurement of students' abilities with respect to the FM-LP using the FCI.

The results are also informative for future research on LPs on the force concept. LPs are hypothetical by nature and, therefore, require empirical validation (Duncan & Hmelo-Silver, 2009). The original work of Alonzo and Steedle (2006) provided an OMC instrument specifically designed to validate FM-LP. Our study shows that the FCI might be used to generate resources to further substantiate the FM-LP's validity. If using the FCI for that purpose, our analyses show that only a subset of 17 items should be used—those matching the LP's content. The consensus assignment of FCI response options to LP level shows that all LP levels are covered among the 17 items. However, LP level 1 is represented by only five item responses. To gain valid inferences on the FM-LP, then we suggest that items should have a closer alignment with the LP and contain responses with more balanced representation of the LP levels.

As the FCI was originally built on a rather comprehensive ontology of the force concept, the remaining 13 FCI items that could not be coded with respect to the FM-LP might serve as a resource to identify further aspects of the force concept. As a next step, LPs addressing these aspects of the force concept would need to be developed and validated. In doing so, LPs on the force concept as a whole could be successively extended through these *aspect LPs*. Furthermore, future studies should investigate how such aspect LPs relate to each other, and whether and how they may constitute a LP of the full force concept.

References

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education, 93*(3), 389-421.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*(5), 765-793.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment 11*(1), 33-63.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching, 46*(6), 606-609.

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: a review and analysis. *Studies in Science Education, 47*(2), 123-182. doi: 10.1080/03057267.2011.604476

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378-382. doi: 10.1037/h0031619

Halloun, I., Hake, R. R., & Mosca, E. P. (1995). Force Concept Inventory Revised Form. *Unpublished Document*. Retrieved from http://modeling.asu.edu/R&E/research.html

Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory. *Physics Teacher, 33*(8), 502.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*, 141-158.

Jung, W., & Wiesner, H. (1981). Verständnisschwierigkeiten beim physikalischen Kraftbegriff. Eine Untersuchung zum Kraftbegriff bei Physikstudenten. *Physik und Didaktik, 9*(2), 111-122

Lee, H.-S., & Liu, O. L. (2010). Assessing learning progression of energy concepts across middle school grades: The knowledge integration perspective. *Science Education, 94*(4), 665-688.

Liang, L. L., Fulmer, G. W., Majerich, D. M., Clevenstine, R., & Howanski, R. (2012). The effects of a model-based physics curriculum program with a Physics First approach: A causal-comparative study. *Journal of Science Education and Technology, 21*(1), 114-124.

Liu, X., & Lesniak, K. M. (2006). Progression in children's understanding of the matter concept from elementary to high school. *Journal of Research in Science Teaching, 43*(3), 320-347. doi: 10.1002/tea.20114

Minstrell, J. (1992). Facets of students' knowledge and rel-

Published on 20/06/2013 on http://edoc.hu-berlin.de/serl

evant instruction. *Paper presented at the Research in Physics Learning: Theoretical Issues and Empirical Studies*, Bremen, Germany.

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching, 46*(6), 675-698. doi: 10.1002/tea.20314

National Research Council [NRC]. (2007). Taking Science to School: Learning and Teaching Science in Grades K-8. Washington, DC: The National Academies Press.

National Research Council [NRC]. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.

Planinic, M., Ivanjek, L., & Susac, A. (2010). Rasch model based analysis of the Force Concept Inventory. *Physical Review Special Topics - Physics Education Research, 6*(1), 010103-010101--010103-010111.

Savinainen, A., & Scott, P. (2002). The Force Concept Inventory: A tool for monitoring student learning. *Physics Education, 37*(1), 45.

Schecker, H., & Niedderer, H. (1986). Das Vorverständnis von S-II-Schülern zur Mechanik. In H. Mikelskis (Ed.), *Zur Didaktik der Physik und Chemie. Vorträge auf der GDCP-Tagung 1985* (181-185). Alsbach: Leuchtturm

Shavelson, R. J., & Kurpius, A. (2012). Reflections on learning progressions. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 13-26). Rotterdam, the Netherlands: Sense Publishers.

Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching, 46*(6), 699-715. doi: 10.1002/tea.20308

Stewart, J., Griffin, H., & Stewart, G. (2007). Context sensitivity in the Force Concept Inventory. *Physical Review Special Topics - Physics Education Research, 3*(1), 010102.

Talanquer, V. (2009). On Cognitive Constraints and Learning Progressions: The case of "structure of matter". International Journal of Science Education, 31(15), 2123-2136. doi: 10.1080/09500690802578025

Wang, J., & Bao, L. (2010). Analyzing Force Concept Inventory with Item Response Theory. *American Journal of Physics, 78*(10), 1064-1070.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. Journal of Research in Science Teaching, 46(6), 716-730. doi: 10.1002/tea.20318

Appendix: Clarification of LP Level Interpretation

The original FM-LP already includes detailed descriptions of students' thinking on each level including typical misconceptions as well; additionally students' typical responses are described for four situations: force, no force, motion and no motion (Alonzo & Steedle, 2009). We used this description – without any changes – as a coding rubric to score the FCI response options. In the following, we give the short LP level descriptions as published by Alonzo and Steedle and illustrate how exemplary FCI response options were representing a respective LP level. Due to the LP being used as a coding rubric, note that many of the instances elaborated here are close to the detailed FM-LP descriptions.

- Level 0

"Way off-track" (p. 405)

We classified those response options that represent a less than rudimentary understanding of force and motion and that contradict Newtonian reasoning and common sense as Level 0. For example, a decreasing of speed even if accelerating forces are acting was coded a Level 0 response.

- Level 1

"Student understands force as a push or pull that may or may not involve motion." (p. 405)

Alonzo and Steedle (2009) elaborate this description saying that "Force is an internal property of objects related to their weight" (p. 403). Response options that only consider gravitational forces as well as those that represent an inconsistent relation between force and motion were therefore coded as Level 1.

- Level 2

"Student believes that motion implies a force in the direction of motion and that nonmotion implies no force. Conversely, student believes that force implies motion in the direction of the force." (p. 404)

This Level also includes an impetus view of force. Response options including forces due to "hits", "pushes" or "pulls" were assigned to this level. Likewise options saying that motion gradually coming to rest due to forces being "used up" were coded as Level 2.

- Level 3

"Student understands that an object is station-

Published on 20/06/2013 on http://edoc.hu-berlin.de/serl

ary either because there are no forces acting on it or because there is no net force acting on it." (p. 403)

Alonzo and Steedle (2009) emphasize that students on this level relate forces to velocity instead of acceleration. Response options showing this conflation, like motions with constant speed even if a force is acting, were therefore assigned to this level.

- Level 4

"Student understands that the net force applied to an object is proportional to its resulting acceleration (change in speed or direction) and that this force may not be in the direction of motion." (p. 403)

All correct responses were assigned Level 4 as they represent a fully developed Newtonian reasoning.

Published on 20/06/2013 on http://edoc.hu-berlin.de/serl